Navy Personnel Research and Development Center
San Diego, California 92152-6800 TR-91-6 March 1991

The P300 Component of the Auditory Event-related Potential: Interlaboratory Consistency and Test-retest Reliability

Leonard J. Trejo
Mark Inlow
Robert R. Stanny
William A. Morey
Scott Makeig
David A. Kobus
Steven A. Hillyard

Released by

T. F. Finley
Captain, U.S. Navy
Commanding Officer
Navy Personnel Research and Development Center

James A. Brady
Captain, U.S. Navy
Commanding Officer
Naval Aerospace Medical Research Laboratory

G. R. Banta
Commander, U.S. Navy
Commanding Officer
Naval Health Research Center

NPRDC-TR-91-6 March 1991

The P300 Component of the Auditory Event-related Potential: Interlaboratory Consistency and Test-retest Reliability

Leonard J. Trejo
Mark Inlow
Navy Personnel Research and Development Center

Robert R. Stanny
William A. Morey
Naval Aerospace Medical Research Laboratory

Scott Makeig
David A. Kobus
Naval Health Research Center

Steven A. Hillyard
University of California, San Diego

Reviewed and approved by

R. C. Sorenson
Technical Director (Acting)
Navy Personnel Research and Development Center

Carl E. Williams
Chief Scientist (Acting)
Naval Aerospace Medical Research Laboratory

John Silva
Scientific Director
Naval Health Research Center

Wigbert C. Wiederholt
Chairman (Acting)
Neurosciences Department
University of California, San Diego

Navy Personnel Research and Development Center
San Diego, California 92152-6800

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATE COVERED
	March 1991	Final--Oct 88-Sep 89

4. TITLE AND SUBTITLE
The P300 Component of the Auditory Event-related Potential: Interlaboratory Consistency and Test-retest Reliability

5. FUNDING NUMBERS
PE 0602234N
RS34H21

6. AUTHOR(S)
L. J. Trejo, M. Inlow, R. R. Stanny, W. A. Morey, S. Makeig, D. A. Kobus, S. A. Hillyard

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Navy Personnel Research and Development Center, San Diego, CA 92152-6800; Naval Aerospace Medical Research Laboratory, Pensacola, FL 32508-5700; Naval Health Research Center, P.O. Box 85122 San Diego, CA 92138-9174; University of California, San Diego, La Jolla, CA 92093

8. PERFORMING ORGANIZATION REPORT NUMBER
NPRDC-TR-91-6

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Technology (Code 222)
800 N. Quincy Street
Arlington, VA 22217

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Also published as NAMRL-1363.

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The auditory evoked potential in a signal classification task using rare and frequent tones was measured by four independent laboratories, to assess the test-retest reliability and interlaboratory consistency of the P300 component. Across laboratories, 61 male subjects (three samples of military subjects, n = 25, n = 18, and n = 8; one sample of college students, n = 10), ranging between 18 and 49 years of age were tested. With few exceptions, all experimental and subject factors were controlled. At each of three electrode sites (Fz, Cz, and Pz), peak amplitude, peak latency, and root-mean-square amplitude (RMS), of the P300 were computed for each subject in test and retest conditions. Statistical analyses of the data supported three strong inferences: (1) test-retest reliability and interlaboratory consistency of P300 measures is high for RMS amplitude, lower for peak amplitude, and lowest for peak latency; (2) recordings from site Pz provide higher reliability and consistency than sites Fz or Cz; (3) the RMS amplitude of P300 is negatively correlated with reaction time to correctly classify rare tones. The data support the notion that ERP components, such as P300, can be reliably measured by different laboratories, with a test-retest correlation coefficient of 0.8 or higher.

14. SUBJECT TERMS
Physiological metrics, psychophysiology, event-related potentials, P300 component, test-retest reliability, interlaboratory consistency

15. NUMBER OF PAGES
15

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNLIMITED

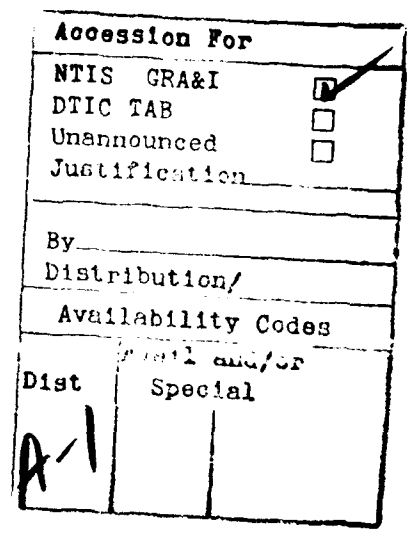
FOREWORD

This report describes experiments performed under the project Biopsychometric Assessment of Combat Operations or BIOPS (PE 0602234N, Task RS34H21), in order to verify procedural uniformity of the laboratories that are performing electrophysiological recording of event-related potentials (ERPs). The experiments involved the presentation of sequences of auditory stimuli (tone bursts of two frequencies) that elicit a positive-going wave in the ERP known as the P300 component. The laboratories were the Navy Personnel Research and Development Center (NPRDC), the Naval Aerospace Medical Research Laboratory (NAMRL), the Naval Health Research Center (NHRC), and the ERP laboratory of the Neurosciences Department at the University of California, San Diego (UCSD). The research was sponsored by the Office of Naval Technology.

The results of the experiments show that, with minor exceptions, the laboratories involved in the BIOPS project have achieved the level of procedural uniformity required to allow for independent advanced investigations in later phases of the project. Furthermore, the results provide data on the reliability and relative value of different measures of the P300 component, which will be useful for future research and development efforts in biopsychometrics and human factors.

T. F. FINLEY
Captain, U.S. Navy
Commanding Officer

RICHARD C. SORENSON
Technical Director (Acting)



SUMMARY

Problem

The demands of modern combat systems have the potential for exceeding the capacity to accurately process information, especially during times of great stress. The capacity of the human to perceive, integrate, remember, and use information may be challenged when the individual is monitoring radar and sonar displays, operating electronic warfare systems, or flying aircraft. Exceeding the capacity of the human operator in such situations may impair decision making and could result in costly tactical errors.

Although much is being done to improve the hardware reliability of combat systems, not enough is being done to improve the performance of system operators. The most unpredictable element in combat systems is often the human operator. Traditional personnel testing and training technologies have not eliminated this unpredictability. In part, this is because traditional methods tend to measure or enhance what a person knows rather than how a person processes information.

The current research is driven by the Navy's need for better methods of assessing the performance of combat system operators, particularly for predicting the ability of operators to continue to make appropriate decisions under heavy workloads, sustained or continuous operations, and in vigilance tasks.

Objective

One class of methods, biopsychometrics, seeks to use physiological data to predict or monitor human performance in operational settings. A range of biopsychometric experiments has been planned in the areas of radar/sonar monitoring, aircrew sustained operations, and pilot performance under g-force stress. Prior to the execution of these experiments, a standardized experiment was performed to ensure procedural uniformity among the laboratories involved. The goal was to ensure consistency and reliability of psychophysiological methods, hardware, and software among the laboratories and future compatibility of the databases to be acquired.

Approach

The standard experiment chosen for this task is known as an "auditory oddball" experiment. In a typical variant of this experiment, event-related potentials (ERPs) are recorded from the scalp while subjects listen to a series of brief tone bursts. A fraction (20%) of these tones differ from the majority in some physical attribute, such as frequency or intensity. Subjects are required to classify tones by pressing a button. Under these conditions, the ERP recorded over the midline of the scalp is characterized by a large positive wave that is maximal at about 300 to 500 milliseconds (ms) after the onset of the rare or "oddball" tones. This component, known as P300, is usually small or absent for the frequent tones.

In this project, the approach was to use this well-known psychophysiological effect as a standard by which the procedures for data acquisition and analysis at the participating laboratories could be calibrated. Within practical limits, all variables that could alter this effect were to be controlled. These included: subject variables such as age, occupation, intelligence, hearing, and

handedness; stimulus variables such as intensity, audio frequency, rise and decay times, and background noise spectrum and intensity; procedural variables such as the number of subjects, probabilities of rare and frequent stimuli, interstimulus interval duration, number of trials per block, number of blocks of trials in test and retest sessions, instructions to subjects, and method of responding; and ERP recording variables such as electrode type, scalp recording sites, reference electrode location, electrooculogram (EOG) recording, subject ground, amplification, calibration signals, analog and digital filtering bandpass, analog digitization rate, and duration of recording epoch for each trial. Finally, analytical procedures were also standardized, including methods of signal averaging, measurement of ERP components, hypothesis testing, and computation of statistics.

Results and Conclusions

Since our primary purpose was to demonstrate interlaboratory consistency and test-retest reliability of the P300 oddball effect, we focused our analysis on the most reliable measure of P300 of several that we tested (including peak amplitude, peak latency, and root-mean-square or RMS amplitude). RMS amplitude was the most reliable measure of P300 both within and between laboratories. This is an integrated area measure of the voltage in the average ERP waveform over a period extending from 275 to 375 ms after the onset of the stimulus. Furthermore, the analysis was focused at recording site Pz (parietal midline electrode referred to average mastoids) since that site exhibited the maximum P300 amplitude and consistency. We first took a logarithmic transformation of the RMS measure to normalize its distribution. Then, using analysis of variance (ANOVA), we tested the factors: laboratory (Navy Personnel Research and Development Center (NPRDC), Naval Aerospace Medical Research Laboratory (NAMRL), Naval Health Research Center (NHRC), the ERP Laboratory of the Neurosciences Department at the University of California San Diego (UCSD)), block (test, retest), and stimulus (rare, frequent) on the dependent measure, log P300 RMS amplitude at site Pz. As expected, only stimulus was significant. No other factors and no interactions were significant.

Four other hypotheses were also tested, using multivariate analysis of variance (MANOVA), to evaluate the effects of electrode sites and ERP measures. These hypotheses concerned (1) the magnitude of the oddball effect on ERP amplitude and latency measures, (2) block-wise stability of the oddball effect on ERP measures, (3) the correlation between choice RT and the magnitude of the oddball effect on ERP measures, and (4) stability of the results across all laboratories. The overall picture they provide is that interlaboratory consistency and test-retest reliabilities of P300 are lower for peak amplitude and latency measures than for RMS measures, and lower at frontal and central recording sites than at the parietal recording site. In addition, correlation analyses showed that the RMS measure of P300 was significantly and negatively correlated with reaction time (RT) for the tone classification response. The reliabilities of P300 measures varied somewhat among the laboratories, possibly due to sampling variation. However, the values of test-retest correlations obtained for the largest sample were consistent with similar correlations reported in the literature.

CONTENTS

LIST OF TABLES

INTRODUCTION

The work reported here is the initial effort of a large project, Biopsychometrics of Combat Operations (BIOPS), that involves psychophysiological experiments in several laboratories (Navy Personnel Research and Development Center (NPRDC), Naval Aerospace Medical Research Laboratory (NAMRL), Naval Health Research Center (NHRC), and the ERP Laboratory of the Neurosciences Department at the University of California San Diego (UCSD)). The purpose of the project is to develop a class of biopsychometric measures, the event-related potentials (ERPs), for use in monitoring cognitive functions of individual military personnel in order to improve the performance of complex systems of which they are a part. These measures will serve as tools for assessing task performance of military personnel and may be used operationally, providing performance-enhancing feedback. ERP measures may also be used in laboratory settings or during training as aids for the design and evaluation of combat systems, work schedules, drug effects, and training systems.

A range of biopsychometric experiments is planned in the areas of radar/sonar monitoring, aircrew sustained operations, and pilot performance under g-force stress. Prior to the execution of these experiments, a standard experiment was performed to ensure procedural uniformity among the laboratories involved. The goal was to ensure interlaboratory consistency and test-retest reliability of ERP recording methods, hardware, and software and future compatibility of the databases to be acquired. This study consists of a standard experiment performed in four laboratories participating in the project.

The standard experiment chosen for this task is known as an "auditory oddball" experiment. In a typical variant of this experiment, ERP signals are recorded from the scalp while subjects listen to a series of brief tone bursts. A fraction of these tones differ from the majority in some physical attribute, such as frequency or intensity. Subjects are required to classify the tones by pressing a button. Under these conditions, the ERP recorded over the central and parietal midline of the scalp is characterized by a large positive wave that is maximal at about 300 to 500 ms after the onset of the rare or "oddball" tones. This component, known as P300, is usually small or absent for the frequent tones. In this experiment, the approach was to use the P300 as a standard by which the procedures for data acquisition and analysis at the laboratories could be calibrated. Within practical limits, all variables that could alter this effect were carefully controlled.

What we now refer to as the P300 component of the event-related potential was originally discovered by Sutton, Braren, Zubin, and John (1965). Recently, Fabiani, Gratton, Karis, and Donchin (1987) reviewed the literature concerning P300, performed experiments to assess methods for its identification and measurement, and estimated its intrasubject reliability. Their study indicated that P300 has five critical defining features: (1) a positive polarity with respect to mastoid or earlobe reference, (2) a latency in excess of 275 ms, (3) a clear peak in the morphology of the waveform, (4) a scalp distribution in which voltage at parietal and central midline electrodes (Pz and Cz) exceeds that at the frontal midline electrode (Fz), and (5) a well-established relationship to experimental manipulations. Our experimental and analytical procedures were patterned after these definitions and our data replicate and extend the findings of Fabiani et al. (1987) to the level of interlaboratory consistency analysis.

METHODS

Standardization of Recording Procedures

A minimal set of common procedures was followed by each laboratory in order to maximize comparability of results. Specifically, the purpose of standardization was to ensure that experimental results obtained from any one laboratory were not due to unique recording procedures.

ERP (or EEG) Electrodes

Three midline electrode sites, Fz, Cz, and Pz, were recorded by each laboratory. In order to reduce variance in electrode placement, NPRDC, NHRC, and UCSD used nylon helmets with fixed tin electrodes mounted in plastic wells arrayed according to the International 10-20 System (Jasper, 1958). NAMRL used individually placed 10-mm gold-cup electrodes for EEG recording. The variance of repeated electrode placements at a single site was estimated to be less than 5 percent (about 1.5 cm). All electrode impedances were checked and kept below 5 kohm at 30 Hz.

ERP Reference

A digitally-derived (average) linked mastoid reference was used by all laboratories. This was derived by a separate recording between the mastoids and subtracting (off line) half the voltage of the active mastoid signal (A2) from all electrodes referred to the reference mastoid (A1).

EOG Recording

Vertical and horizontal electrooculograms (EOGs) were recorded using surface-mounted electrodes (Ag-AgCl at NPRDC, NHRC, and UCSD; gold at NAMRL). Impedances were kept below 10 kohm at 30 Hz. Vertical EOG was recorded as the voltage difference between an electrode placed 1 cm above the superior rim of the right eye orbit and an electrode placed at the inferior rim of the right eye, along a vertical line intersecting the pupil. Horizontal EOG was recorded as the voltage difference between an electrode placed 2 cm laterally to the outer canthus of the right eye and an electrode placed 2 cm laterally to the outer canthus of the left eye, along a line intersecting both pupils.

Subject Ground

Subject ground was on the midline at a point 3 cm anterior to the Fz electrode. Ground impedance was kept below 10 kohm at 30 Hz.

Digitization

Analog ERP data were digitized at a minimal sampling rate of 125 Hz. Overall gain and analog-to-digital (A/D) conversion provided a dynamic range of $\pm$250 microvolts (μv) and a maximal quantization step size of 0.15 μv. All single-epoch ERP data were stored permanently on magnetic mass-storage media.

EOG Amplification

Since EOG signals were to be used to correct ERP recordings for ocular artifact, it was important to ensure adequate EOG amplification and dynamic range. For this reason, EOG amplifier gain was adjusted to be not less than 1/5 the gain of the ERP amplifiers, allowing for a minimal dynamic range of ± 250 μv and a maximal quantization step size of 0.75 μv.

Analog Bandpass

Analog filters were linear and free from significant phase distortion within the bandwidth of the ERP and EOG signals (0.1 to 100 Hz).

Calibration

All signal amplifiers were calibrated individually using a fixed voltage signal source. Calibration was performed at least once before or after each subject.

Digital Filtering

ERP data were digitally filtered off-line using zero-phase-shift filters to reduce high frequency noise in the band between 30 and 100 Hz.

Recording Period

The recording period for each single ERP included a minimum of a 100-ms pre stimulus interval and a 900-ms post-stimulus interval.

Standardization of the Auditory Oddball Task

Frequent Stimulus

The frequent stimulus was a pure tone burst (sine wave) with a frequency of 1500 Hz and a duration of 50 ms. During the first 10 ms, the amplitude of the stimulus rose linearly from zero to maximum and, during the last 10 ms, the amplitude fell linearly from maximum to zero. The frequent stimuli occurred with a probability of 0.8.

Rare Stimulus

The rare stimulus was a pure tone burst (sine wave) with a frequency of 750 Hz, duration of 50 ms, and probability of occurrence of 0.2. During the first 10 ms, the amplitude of the stimulus rose linearly from zero to maximum and, during the last 10 ms, the amplitude fell linearly from maximum to zero.

Stimulus Intensity

The intensity of both rare and frequent stimuli was 70 dB. At NAMRL, NHRC, and UCSD, intensity was set relative to sensation level using calibrated attenuators. NPRDC measured the

intensity relative to sound pressure level with an impulse sound level meter (using the "A" weighting, which compensates for the differential frequency sensitivity of the ear).

Background Noise

Background noise was controlled with a white noise generator and had an intensity of 60 dB relative to sensation level. Intensity was measured with a sound level meter. NHRC made no provisions for controlled background noise.

Stimulus Presentation

Stimuli were presented binaurally with headphones. All headphones had a linear frequency response over the range required to present the stimuli (750 and 1500 Hz tone bursts) and did not produce electrical recording artifacts. Subjects were seated upright in an armchair with eyes open, viewing a fixation point in an unstructured visual field.

Interstimulus Interval

The interval between stimuli varied randomly from 1 and 1.5 seconds and was unpredictable by the subjects within that range.

Number of Trials

A minimum of 200 trials (frequent + rare) was collected for each subject per block.

Number of Blocks

At least two blocks per subject, repeated under standard conditions within a single session, were performed to allow for estimates of test-retest reliability (see below). At NPRDC, 12 subjects received an additional test session (two more blocks) approximately one week after the first session.

Subjects

Subjects were interviewed prior to testing to make sure that they were well rested, alert, willing to participate, and not under the influence of any medications, including tobacco, caffeine, antihistamines, analgesics, sedatives, narcotics, antidepressants, stimulants, alcohol, or prescription drugs. Subjects were assessed for these variables with a questionnaire.

Additional a priori subjects' specifications included the following: (1) A minimum of 20 military subjects per laboratory was requested; (2) all subjects were required to have Armed Forces Qualifications Test (AFQT) scores within mental categories I and II, equivalent IQ scores, or other evidence of above-average mental ability; (3) all subjects were required to be males between the ages of 18 and 30 years; (4) each subject's hearing was to be checked or verified from recent (6 months) medical records; and (5) subjects were required to be right handed as assessed by self-report.

Only the hearing tests, number of subjects tested, and age requirements were not strictly adhered to in the study. At NPRDC, hearing was assessed by self-report as opposed to medical records or audiogram. The number of subjects who completed the experiment at each laboratory was: 25 at NPRDC, 18 at NAMRL, 8 at NHRC, and 10 at UCSD. At NHRC, 3 of the 8 subjects tested exceeded the age specifications (35, 36, and 49 yrs). UCSD did not use military subjects; college students served instead.

Average ERPs

At NPRDC and NAMRL, the data for each subject consisted of average ERPs computed from a minimum of 25 rare stimulus trials in each block and from an equal number of randomly selected frequent trials for a total of four average ERPs per subject. At NHRC and UCSD, all of the frequent stimulus trials (typically about 200) were included in the average ERPs for frequent stimuli. Minimum signal bandpass for average ERPs after filtering was 0.3 to 25 Hz.

Component Peak Analyses

From each average ERP, two oddball-related component peak measures were computed: P300 component peak amplitude (pre-stimulus average baseline to peak) and P300 peak latency relative to stimulus onset. P300 was operationally defined for the oddball task as the maximal positive peak between 275 and 425 ms post-stimulus at site Pz referenced to average mastoids.

The N1 and P2 component measures, baseline-to-peak amplitude and peak latency, were also computed. N1 was operationally defined as the maximal negativity relative to pre-stimulus average baseline between 80 and 140 ms at site Cz; P2, as the maximal positivity relative to pre-stimulus average baseline between 140 and 200 ms at site Cz.

The sampling rates used by the laboratories did not always provide samples at the exact boundaries of the intervals for the component peak analyses. In these cases, the sample following the defined interval boundary was used. For example, in the NPRDC data, there was no sample at 425 ms post-stimulus, the upper boundary of the P300 interval. In this case, the sample at 429.7 ms was used as the upper boundary of the P300 interval. Similar boundary variations appeared in the intervals of other laboratories and components (see the Appendix). However, none of these variations exceeded the desired window boundaries by more than 5 ms.

Root-mean-square (RMS) Amplitude Analysis

From the frequent and rare average ERPs (see above), RMS amplitudes in a single 100-ms time interval between 275 and 375 ms post-stimulus were computed according to the method described by Trejo (1988) with one exception: Average ERP waveforms were adjusted to have a zero-mean pre-stimulus epoch instead of a zero mean for the entire averaging epoch.

Behavioral Responses

Subjects were instructed to fixate on a dot on a wall or a video monitor, listen to the tones, and press a "target" button for rare tones or a "non-target" button for frequent tones on each trial. They were also instructed to respond as quickly as possible without sacrificing accuracy. At NPRDC,

NAMRL, and NHRC, subjects responded by pressing one button on a two-button response panel using the right middle finger (non-target) or right index finger (target). At UCSD, subjects used a pair of joysticks with thumb switches for the left (non-target) and right (target) hands. The response panels did not require gross motor activity other than finger/wrist movements. Subjects were given instructions and a minimum of 100 trials of practice using one button to signal detection of high-pitched tones and the other for low-pitched tones. Choice reaction time (RT) was measured with an accuracy of $\pm$ 4 ms relative to stimulus onset for each trial.

Hypothesis Testing

The oddball effect was defined as the difference in ERP dependent measures between average ERPs for rare and frequent stimuli. The three oddball-related dependent measures were P300 amplitude (operationally defined as the maximum positive peak between 275 and 425 ms post-stimulus), latency of the peak P300 amplitude (between 275 and 425 ms post-stimulus), and RMS amplitude between 275 and 375 ms post-stimulus. Hypotheses to be tested concerned (1) the magnitude of the oddball effect on ERP amplitude and latency measures, (2) block-wise stability of the oddball effect on ERP measures, (3) the correlation between choice RT and the magnitude of the oddball effect on ERP measures, and (4) stability of the results across all laboratories.

To test these hypotheses, both multivariate analysis of variance (MANOVA) and correlation analysis were performed. Separate MANOVAs were performed for P300 amplitude, P300 latency, and RMS amplitude measures. In these MANOVAs, the factors included (1) laboratory, (2) stimulus, (3) electrode site, and (4) block (the test-retest factor).

Specific null hypotheses addressed using analysis of variance (ANOVA) on single dependent measures were: (1) The sample mean (e.g., of P300 amplitude, P300 latency, or RMS amplitude) is equal in average ERPs for different levels of each factor and (2), with respect to the sample mean, factors do not interact.

Specific null hypotheses addressed using correlation analysis were: (1) The magnitude of the oddball effect on dependent measures (e.g., of P300 amplitude, P300 latency, or RMS amplitude) in block 1 is linearly independent of the magnitude of the same effect in block 2 (reliability as assessed with bivariate correlation); (2) the average magnitude of the oddball effect on a dependent measure is linearly independent of average RT to rare stimuli (performance-ERP relationship assessed with bivariate correlation); and (3) the magnitude of the oddball effect on a dependent measure is linearly independent of average RT to frequent stimuli (bivariate correlation). Non-oddball related measures included amplitude and latency of N1 and P2. Specific hypotheses for these measures were not stated. However, summary tables of the means and standard deviations of these measures, and their correlations with reaction time, percent correct, and age variables across blocks and sites appear in the Appendix, along with P300 summary tables. P2 peak amplitude measures were not reported by NHRC and UCSD.

Artifact Processing

The laboratories adopted a minimum set of criteria for the acceptance of data as being free from artifacts, including electrical artifacts from extraneous sources, digital artifacts (e.g., overflow or underflow, truncation, etc.), and non-cephalic bioelectric artifacts (electrooculographic,

electromyographic, electrocardiographic, motion, skin conductance changes, etc.). With some exceptions, the following criteria were observed by each laboratory.

Transient Artifacts

All single trials for which the voltage difference between the active electrode and the reference contained a transient (<100 ms in duration) signal with peak-to-peak amplitude of 50 μv or more were either rejected from analysis or corrected if the artifact source could be independently estimated. Cancellation techniques reduced the artifact amplitude by a factor of at least 20 dB. Epochs containing any transient artifact which saturated input amplifiers or produced numerical overflow or underflow of computer registers were rejected. For ocular artifacts, NPRDC and NAMRL applied the correction procedure developed by Gratton, Coles, and Donchin (1983) to the ERP data. NHRC and UCSD rejected any trials containing ocular artifacts.

Periodic Artifacts

All single epochs which contained identified non-cephalic periodic artifacts (e.g., 60-Hz AC line noise) with peak-to-peak amplitude exceeding 10 μv were rejected or corrected. Either analog notch or comb filters tuned to the offending frequencies or a suitable digital filter were employed to reduce the peak-to-peak amplitude to less than 10 μv.

RESULTS AND CONCLUSIONS

All laboratories provided three P300 measures at sites Fz, Pz, and Cz: peak amplitude measured from the average pre-stimulus baseline voltage (referred to as "amplitude"), RMS amplitude (referred to as "RMS"), and peak latency (referred to as "latency") (see Methods). Grand average ERP data from one laboratory (NPRDC) for rare and frequent stimuli, and the standardized measurement time windows are shown in Figure 1. Grand average ERP data for all laboratories, separated by blocks and stimulus type are shown in Figure 2. Positive polarity of ERP voltage is up in both figures.

P300 measures for all laboratories and electrode sites are listed in Table 1. In support of our operational definition of P300, we found that, for rare stimuli, P300 amplitude at site Pz had the greatest amplitude and lowest coefficient of variation (standard deviation/mean) of all electrode sites. For this reason, we used the P300 amplitude measures at site Pz for our single-site analyses.

Behavioral data reported by each laboratory for all subjects included average RT for rare and frequent stimuli, and percent correct classification of stimuli. For response purposes, correct classifications were called "hits" for rare stimuli and "correct rejections" for frequent stimuli. Mean percent correct, RTs for hits and correct rejections, and associated standard deviations are listed by laboratory and block in Table 2.

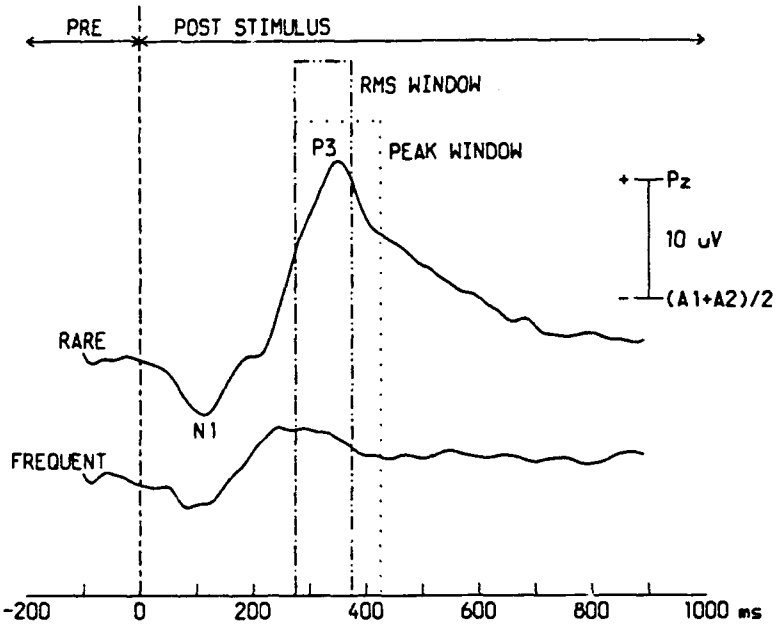


Figure 1. Grand average ERPs for rare and frequent recorded at site Pz for one laboratory (NPRDC). Also shown are the time windows used for computing the P300 RMS value and for determining P300 amplitude and latency.

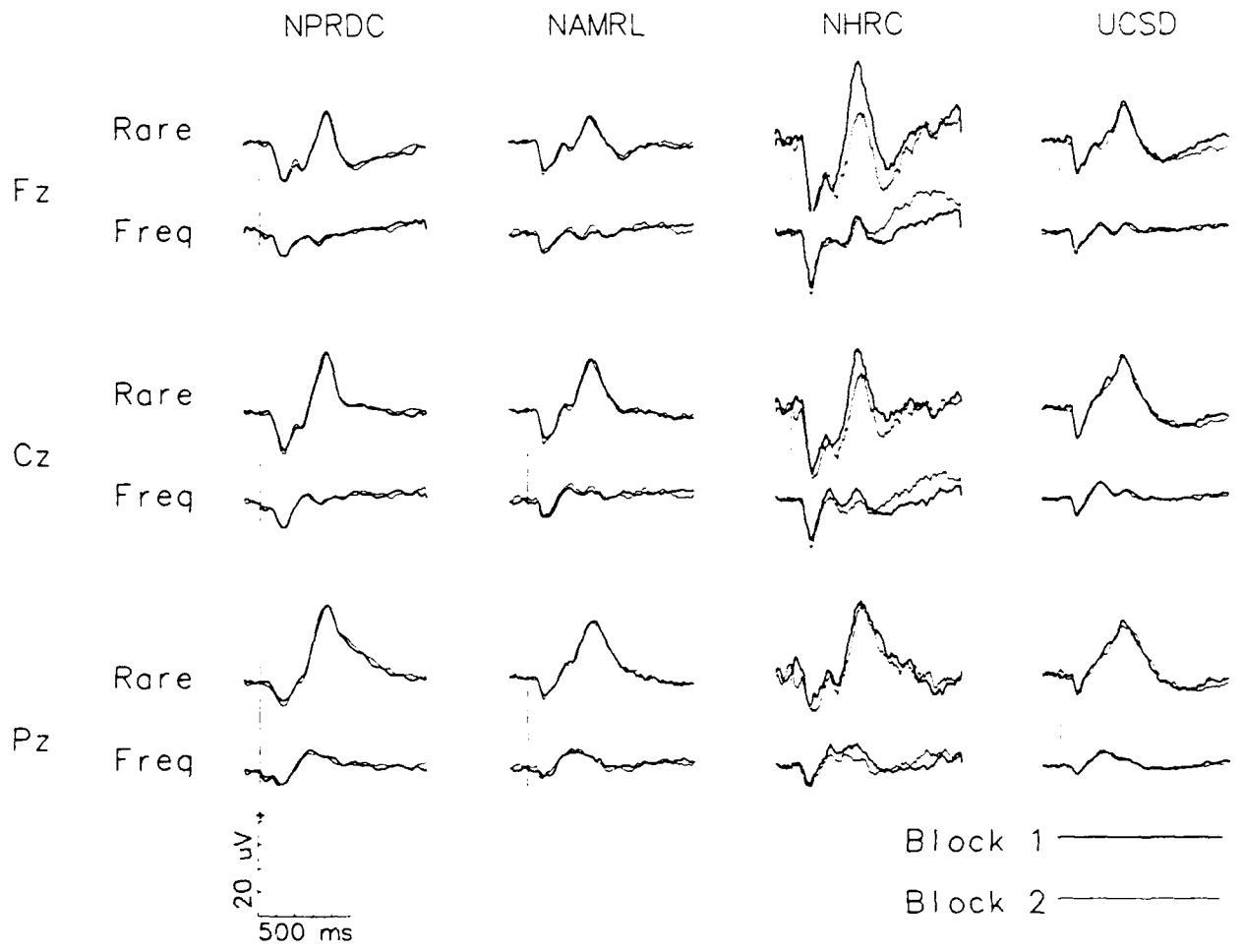


Figure 2. Grand average ERP data for all laboratories by block and stimulus type.

Table 1

P300 Peak Amplitudes for Rare and Frequent Stimuli by Site and Lab

Lab	Block	Fz		Cz		Pz	
		Mean	CV	Mean	CV	Mean	CV
Rare Stimuli							
Pooled	1	9.0	0.67	13.8	0.53	16.3	0.44
	2	7.3	0.88	12.7	0.58	15.8	0.44
NPRDC	1	7.5	0.88	15.0	0.57	19.6	0.44
	2	6.4	0.95	14.3	0.63	19.5	0.45
NAMRL	1	6.9	0.46	11.9	0.49	12.6	0.47
	2	6.4	0.66	11.8	0.45	12.6	0.34
NHRC	1	16.8	0.46	13.4	0.49	16.8	0.43
	2	10.0	1.12	9.9	0.75	16.4	0.39
UCSD	1	10.5	0.67	14.5	0.44	14.3	0.34
	2	9.2	0.63	12.3	0.48	12.0	0.50
Frequent Stimuli							
Pooled	1	7.2	0.59	10.5	0.49	12.2	0.47
	2	6.7	0.58	10.0	0.53	11.5	0.52
NPRDC	1	6.6	0.50	11.1	0.50	14.0	0.46
	2	6.2	0.47	10.9	0.50	13.4	0.52
NAMRL	1	5.0	0.49	9.7	0.54	10.8	0.49
	2	5.4	0.53	9.8	0.55	10.8	0.47
NHRC	1	12.8	0.28	10.0	0.41	11.6	0.51
	2	11.2	0.42	8.3	0.49	10.3	0.44
UCSD	1	8.2	0.66	11.1	0.48	10.7	0.39
	2	6.6	0.72	9.3	0.61	9.0	0.58

CV = coefficient of variation (standard deviation/mean).

Table 2

Mean Percent Correct and Reaction Times (in ms)

	Percent Correct		Reaction Times			
			Hits		CR	
Lab	Block 1	Block 2	Block 1	Block 2	Block 1	Block 2
NPRDC	97± 4	96± 5	419± 63	432± 70	344± 54	351± 59
NAMRL	97± 2	97± 3	402± 69	403+ 85	335± 56	330± 60
NHRC	98± 2	98± 2	422± 63	415± 5	366± 82	323± 32
UCSD	88± 13	87± 12	477± 96	487± 90	373± 76	369± 85

Hit = correct classification of rare stimuli.
CR = correct classification of frequent stimuli.

Analysis of Behavioral Responses

We ran a repeated measures ANOVA on RT with laboratory as the between factor, and block and stimulus as the within factors. The results are shown in Table 3. Laboratory and all factor-by-laboratory interactions were not significant. RT was significantly lower for correct rejections than for hits. There was also a significant block-by-stimulus interaction. Table 4 gives the results of simple effects ANOVAs for blocks 1 and 2, and shows that stimulus was independently significant for both blocks. In simple effects ANOVAs for the two stimuli, block was not significant for rare [$F(1, 57) = 0.88$, $p < 0.3521$] or frequent stimuli [$F(1, 57) = 3.07$, $p < 0.0727$].

We then analyzed the percent-correct scores with laboratory as the between factor and block as the within factor. From Table 5, it is clear that laboratory was significant but block and block-by-laboratory were not. UCSD was significantly different from the other laboratories at the $p = 0.01$ level using standard post-hoc pairwise comparison methods, here the Bonferroni, Tukey, and Scheffe methods. The other laboratories did not differ significantly from each other. When UCSD was not included, laboratory was not significant [$F(2, 48) = 0.64$, $p < 0.5337$]; the other results were unchanged. Therefore, we conclude that the laboratories were comparable in their behavioral data with the exception of the lower mean percent correct for UCSD.

Table 3

ANOVA Summary Table for Reaction Times

Source	SS	df	F	p
Lab (L)	90928	3	2.02	0.1217
Error	856334	57		
Block (B)	605	1	0.56	0.4557
B X L	7623	3	2.37	0.0800
Error	61112	57		
Stimulus (S)	346267	1	157.82	0.0000
S X L	11726	3	1.78	0.1609
Error	125058	57		
B X S	2991	1	6.40	0.0142
B X S X L	1508	3	1.08	0.3668
Error	26643	57		

Table 4

Simple Effects ANOVA Summaries for Reaction Times

Source	SS	df	F	p
Block 1				
Lab (L)	43395	3	1.96	0.1310
Error	421721	57		
Stimulus (S)	142448	1	93.06	0.0000
S X L	6257	3	1.36	0.2634
Error	87252	57		
Block 2				
Lab(L)	55155	3	2.11	0.1085
Error	495725			
Stimulus (S)	206810	1	182.91	0.0000
S X L	6977	3	2.06	0.1161
Error	64448			

Table 5

ANOVA Summary Table for Percent Correct

Source	SS	df	F	p
Lab (L)	0.155111	3	8.36	0.0001
Error	0.352710	57		
Block (B)	0.000980	1	1.32	0.2561
B X L	0.000571	3	0.26	0.8569
Error	0.042450	57		

Analysis of P300 Baseline-to-peak Amplitude

To test the hypotheses outlined in the Methods section, we ran repeated measures MANOVAs on the various P300 measures, with each multivariate observation consisting of the measure's value at the three scalp sites. First, we ran tests to verify assumptions of multivariate normality of the observations for small sample size and equal covariance across the various populations. From histograms and Shapiro-Wilks univariate normality tests on the individual components, the assumption of multivariate normality appeared reasonable for the multivariate P300 amplitude measure. However, the data forced rejection of the assumption of equal covariance across laboratories for block 1 ($p < 0.0001$) and block 2 ($p < 0.0001$).

We then performed a three-way repeated measures MANOVA with laboratory as the between factor, and block and stimulus as the within factors, with each multivariate observation consisting of the P300 amplitude at site Fz, Cz, and Pz. The results are summarized in Table 6. The main effect for laboratory (hereafter referred to as "laboratory") was significant. Two interactions were also significant: stimulus-by-laboratory and block-by-stimulus. A site-by-site analysis, not shown in Table 6, revealed that the stimulus-by-laboratory interaction was significant only at site Pz [$F(3, 57) = 5.89$, $p < 0.0014$], whereas the block-by-stimulus interaction was significant only at site Fz [$F(1, 57) = 7.82$, $p < 0.007$].

From Figure 2 it appears that the block-by-stimulus interaction is due to the large difference between block 1 and block 2 rare P300 amplitudes at site Fz for NHRC. We tested this hypothesis by computing a paired t-test on NHRC's site Fz rare P300 amplitudes, block 1 vs. block 2, which was not significant [$t(7) = 1.77$, $p = 0.1205$]. Since the t-test is sensitive to large individual differences, we ran a sign test which gave a p-value of 0.035. We also ran this t-test on the data from the other laboratories. None of these t-tests were significant.

Table 6

MANOVA Summary for P300 Amplitudes

Source	Statistic	F	df_N	df_D	p
Lab (L)	0.384493[a]	7.16	9	134.01	0.0000
Block (B)	7.91782[b]	2.55	3	55	0.0653
B X L	0.808990[a]	1.35	9	134.01	0.2148
Stimulus (S)	280.395[b]	90.19	3	55	0.0000
S X L	0.508267[a]	4.77	9	134.01	0.0000
B X S	8.83668[b]	2.84	3	55	0.0461
B X S X L	0.767088[a]	1.71	9	134.01	0.0916

[a]L-ratio (Wilk's lambda likelihood ratio statistic; Dixon, 1985, p. 395).
[b]T-square (Hotelling's generalized T-zero squared statistic; Dixon, 1985, p. 395).

As shown in Table 7, when NHRC was excluded from the analysis, the block-by-stimulus interaction was no longer significant. However, since stimulus-by-laboratory was still significant at Pz [$F(2, 50) = 8.83$, $p < 0.0005$], simple effects MANOVAs were run for rare and frequent stimuli to test if laboratory was significant for both stimuli. As the results in Table 8 show, laboratory was significant for both rare and frequent stimuli. The block factor and the block-by-laboratory interaction were not significant for either stimulus. Since the sample sizes for the individual laboratories were small, we ran univariate ANOVAs for each laboratory on site Pz to test for significance of stimulus. Stimulus was significant for each laboratory, with the least significant result being $F(1, 7) = 45.35$, $p < 0.0003$.

Since we chose the P300 amplitude at site Pz as one of the estimators for predicting behavioral responses, we ran a three-way repeated measures ANOVA with laboratory as the between factor, and with stimulus and block as the within factors. All laboratories were included. Since there was a stimulus-by-laboratory interaction [$F(3, 57) = 5.89$, $p < 0.0014$], we ran simple effects ANOVAs for both rare and frequent stimuli. There was no difference between laboratories for the frequent stimuli [$F(3, 57) = 0.57$, $p < 0.6377$], but there was a significant difference for the rare stimuli [$F(3, 57) = 4.36$, $p < 0.0078$]. The block factor and the block-by-laboratory interaction were not significant for either stimulus. These results show that the peak amplitude of the P300 for rare stimuli may vary with the sample population and/or with subtle differences in recording or data-processing procedures.

Table 7

MANOVA Summary for P300 Amplitudes (NHRC excluded)

Source	Statistic	F	df_N	df_D	p
Lab (L)	0.573104[a]	5.14	6	96	0.0001
Block (B)	0.949164[b]	0.30	3	48	0.8225
B X L	0.912893[a]	0.75	6	96	0.6140
Stimulus (S)	233.730[b]	74.79	3	48	0.0000
S X L	0.640369[a]	3.99	6	96	0.0013
B X S	4.59165[b]	1.47	3	48	0.2347
B X S X L	0.940834[a]	0.50	6	96	0.8104

[a]L-ratio (Wilk's lambda likelihood ratio statistic; Dixon, 1985, p. 395).
[b]T-square (Hotelling's generalized T-zero squared statistic; Dixon, 1985, p. 395).

Table 8

Simple Effects MANOVA Summaries for P300 Amplitudes

Source	Statistic	F	df_N	df_D	p
Rare Stimuli					
Lab (L)	0.583604[a]	4.94	6	96	0.0002
Block (B)	3.80983[b]	1.22	3	48	0.3129
B X L	0.919268[a]	0.69	6	96	0.6599
Frequent Stimuli					
Lab (L)	0.621041[a]	4.30	6	96	0.0007
Block (B)	1.20304[b]	0.38	3	48	0.7643
B X L	0.928971[a]	0.60	6	96	0.7293

[a]L-ratio (Wilk's lambda likelihood ratio statistic; Dixon, 1985, p. 395).
[b]T-square (Hotelling's generalized T-zero squared statistic; Dixon, 1985, p. 395).

Analysis of P300 RMS Amplitude

For the P300 RMS measure, we performed similar analyses as performed for P300 (baseline-to-peak) amplitude. Unlike the amplitude data, it was necessary to transform the RMS data to distribute it normally. Since the measures tended to be right-skewed, we used the natural log transformation which resulted in a closer approximation to the normal distribution. The test for unequal covariance across laboratories was not significant for block 1 ($p = 0.08$), but was significant for block 2 ($p = 0.0319$). We ran a 3-way repeated measures MANOVA on the RMS with the same design as for the amplitude data: Laboratory was the between factor, and block and stimulus were the within factors with each multivariate observation consisting of the values at sites Fz, Cz, and Pz.

Although the covariance matrices were statistically different across laboratories for block 2, the differences were probably not large enough to invalidate the p-values of the MANOVA for the laboratory effects. The effect of laboratory was significant but a site-by-site analysis (see Table 9) showed that this was true only at site Fz [$F(3, 57) = 12.09$, $p < 0.0001$]. Standard post-hoc multiple comparison procedures showed that NHRC data differed from NPRDC and NAMRL data ($p < .05$), but not UCSD data. None of the other laboratories differed significantly from each other. When NHRC data were excluded, the effect of laboratory was not significant [$F(6, 96) = 1.34$, $p < 0.2457$]. There were no significant interactions between laboratory and the other factors. The block factor and the block-by-stimulus interaction were not significant. The main effect for stimulus was significant.

Table 9

MANOVA Summary for P300 RMS Amplitudes

Source	Statistic	F	df_N	df_D	p
Lab (L)	0.535259[a]	4.36	9	134.01	0.0001
Block (B)	3.97144[b]	1.28	3	55	0.2912
B X L	0.865935[a]	0.91	9	134.01	0.5211
Stimulus (S)	339.946[b]	109.34	3	55	0.0000
S X L	0.852058[a]	1.01	9	134.01	0.4333
B X S	7.42792[b]	2.39	3	55	0.0786
B X S X L	0.884500[a]	0.77	9	134.01	0.6442

[a]L-ratio (Wilk's lambda likelihood ratio statistic; Dixon, 1985, p. 395).
[b]T-square (Hotelling's generalized T-zero squared statistic; Dixon, 1985, p. 395).

Since our primary purpose was to demonstrate consistency and reliability of the P300 oddball effect, we focused the analysis on RMS at site Pz, which was clearly the best site to measure P300 by criteria of largest amplitude and lowest variability (Table 1). We ran an ANOVA on the log P300 RMS amplitude at site Pz with laboratory as the between factor, and block and stimulus as the within factors; results are shown in Table 10. As expected, only stimulus was significant [$F(1, 57) = 285.88$, $p < 0.0000$]. No other factors and no interactions were significant.

Table 10

ANOVA Summary Table for Pz P300 RMS Amplitudes

Source	SS	df	F	p
Lab (L)	3.49	3	1.19	0.3216
Error	55.75	57		
Block (B)	0.47	1	3.04	0.0869
B X L	0.13	3	0.29	0.8335
Error	8.78	57		
Stimulus (S)	80.48	1	285.88	0.0000
S X L	1.01	3	1.19	0.3199
Error	16.05	57		
B X S	0.00	1	0.00	0.9838
B X S X L	0.26	3	0.55	0.6499
Error	9.13	57		

Analyses of P300 Latency

Our final analyses of variance pertain to P300 latency. As with the amplitude measures, we ran a three-way MANOVA on the P300 latencies. The results are summarized in Table 11. Laboratory was not significant, but there was a marginally significant block-by-stimulus interaction.[1] This interaction was independently present for all laboratories because block-by-stimulus-by-laboratory was not significant. That is, block-by-stimulus-by-laboratory would need to be significant in order for block-by-stimulus to be significant for some, but not all, laboratories. For this reason, we ran simple effects MANOVAs for rare and frequent stimuli, and blocks 1 and 2. The results are

[1]For two reasons, the p-values for laboratory and laboratory interactions may not be exact. First, the distribution of P300 latency was not clearly normal in all cases. At some of the laboratory/site combinations, the distributions appeared normal, whereas at others, the distributions appeared bimodal. Given the small sample sizes of each distribution, we decided to accept the assumption of normality. Second, the covariance matrices appeared to differ among the laboratories. Tests for unequal covariances between laboratories were significant for blocks 1 and 2 with p-values of 0.0001 and 0.0001 respectively.

summarized in Table 12. Block and block-by-laboratory were not significant for rare or frequent stimuli. The simple effects MANOVAs for each block indicate that stimulus was significant for both blocks 1 and 2 and that stimulus-by-laboratory was not.

We also ran an ANOVA on the P300 latencies at site Pz. The results paralleled the multivariate analysis except that for frequent stimuli, block was significant [$F(1, 57) = 5.7$, $p < 0.0203$]; as in the multivariate measure, block-by-laboratory was not [$F(3, 57) = 2.34$, $p < 0.0829$].

Table 11

MANOVA Summary for P300 Latencies

Source	Statistic	F	df_N	df_D	p
Lab (L)	0.802260[a]	1.41	9	134.01	0.1894
Block (B)	3.13842[b]	1.01	3	55	0.3956
B X L	0.846311[a]	1.06	9	134.01	0.3990
Stimulus (S)	90.3561[b]	29.06	3	55	0.0000
S X L	0.868295[a]	0.89	9	134.01	0.5365
B X S	8.81155[b]	2.83	3	55	0.0465
B X S X L	0.901860[a]	0.65	9	134.01	0.7563

[a]L-ratio (Wilk's lambda likelihood ratio statistic; Dixon, 1985, p. 395).
[b]T-square (Hotelling's generalized T-zero squared statistic; Dixon, 1985, p. 395).

Table 12

Simple Effects MANOVA Summaries for P300 Latencies

Source	Statistic	F	df_N	df_D	p
Rare Stimuli					
Lab (L)	0.760141[a]	1.78	9	134.01	0.0784
Block (B)	1.26338[b]	0.41	3	55	0.7490
B X L	0.897439[a]	0.68	9	134.01	0.7286
Frequent Stimuli					
Lab (L)	0.893089[a]	0.71	9	134.01	0.7000
Block (B)	6.60513[b]	2.12	3	55	0.1076
B X L	0.864522[a]	0.92	9	134.01	0.5119
Block 1					
Lab (L)	0.84058[a]	1.10	9	134.01	0.3658
Stimulus (S)	42.7437[b]	13.75	3	55	0.0000
S X L	0.859064[a]	0.96	9	134.01	0.4769
Block 2					
Lab (L)	0.819229[a]	1.27	9	134.01	0.2582
Stimulus (S)	91.7806[b]	29.52	3	55	0.0000
S X L	0.920251[a]	0.52	9	134.01	0.8601

[a]L-ratio (Wilk's lambda likelihood ratio statistic; Dixon, 1985, p. 395).
[b]T-square (Hotelling's generalized T-zero squared statistic; Dixon, 1985, p. 395).

Oddball Effect: Reliability Analyses

In order to further assess reliability of the oddball effect, correlations from block 1 to block 2 of the oddball effect on the three P300 measures at site Pz were computed. Values of these correlations for each laboratory, with pooled values where appropriate, are shown in Table 13. We rejected the hypothesis of linear independence for P300 RMS and latency since the pooled correlations were significant at $p < 0.001$. We also rejected linear independence for the peak amplitude since NPRDC and NAMRL had significant correlations at $p < 0.001$ and $p < 0.05$, and NHRC and UCSD had high test-retest correlations also. The correlations for NHRC and UCSD, though large, were not significant, possibly due to the small sample sizes. In summary, the correlations for the amplitude measures were consistently high across the laboratories whereas the latency correlations were more variable.

Table 13

Oddball Effect Test-retest Correlations at Site Pz

Lab	Peak Amplitude[a]	Peak Latency	RMS Amplitude	df
Pooled	---	0.42***	0.66***	59
NPRDC	0.81***	0.58**	0.75***	23
NAMRL	0.49*	0.26	0.50*	16
NHRC	0.59	0.10	0.76*	6
UCSD	0.63	0.48	0.73*	8

[a]Peak amplitude pooled correlations were not computed due to a significant effect of laboratory (see Table 8).
*p < .05.
**p < .01.
***p < .001.

Relationship of P300 to Reaction Time Data

We also used correlation analysis to test for linear relationships between the oddball effect on the P300 measures at site Pz and average RT data. Tables 14, 15, and 16 give the correlation coefficients from all laboratories for each measure and, where appropriate, the correlation coefficients pooled across laboratories. Each table lists the correlation coefficients between the P300 measure of interest and the average RT for hits and correct rejections. Correlation coefficients were computed for each block. The only oddball effect which correlated significantly with RT was the P300 RMS amplitude.[2] For the data pooled across laboratories, the P300 RMS amplitude had a significant negative correlation with both average RT to hits and average RT to correct rejections, and this negative correlation held for both blocks. The correlations within each laboratory were in accord with the pooled correlations. With the exception of the NPRDC data, most individual laboratory correlations were not significant, possibly due to small sample sizes.

[2]For simple component measures, not defined as an "oddball effect," we observed several significant correlations with performance variables RT and percent correct as well as with age (see the Appendix). These findings will be fully described in a future report.

Table 14

Correlations of Peak Amplitude Oddball Effect with Mean Reaction Time

	Block 1		Block 2		
Lab	Hits	CR	Hits	CR	df
NPRDC	-0.04	-0.28	-0.25	-0.23	23
NAMRL	-0.30	-0.26	-0.05	-0.31	16
NHRC	-0.48	-0.43	-0.20	0.22	6
UCSD	-0.18	-0.21	-0.31	-0.13	8

Hit = correct classification of rare stimuli.
CR = correct classification of frequent stimuli.

Table 15

Correlations of RMS Amplitude Oddball Effect with Mean Reaction Time

	Block 1		Block 2		
Lab	Hits	CR	Hits	CR	df
Pooled	-0.30*	-0.36**	-0.38**	-0.39**	59
NPRDC	-0.29	-0.43*	-0.58**	-0.50*	23
NAMRL	-0.34	-0.26	-0.29	-0.55*	16
NHRC	-0.46	-0.55	-0.16	-0.19	6
UCSD	-0.44	-0.23	-0.42	-0.20	8

*$p < .05$.
**$p < .01$.

Table 16

Correlations of Peak Latency Oddball Effect with Mean Reaction Time

	Block 1		Block 2		
Lab	Hits	CR	Hits	CR	df
Pooled	0.04	-0.05	0.22	0.05	59
NPRDC	0.25	0.10	0.25	0.16	23
NAMRL	-0.13	-0.03	0.57*	0.39	16
NHRC	-0.18	-0.05	0.01	-0.32	6
UCSD	0.27	-0.20	-0.08	-0.55	8

*$p < .05$.

DISCUSSION

The goal of this study was to demonstrate test-retest and interlaboratory consistency and reliability of a well-known psychophysiological effect, the P300 component of the ERP for auditory oddball stimuli. Our assumption that P300 is maximal at site Pz was supported by maximum amplitude and minimum variability criteria, and agrees with the findings of others (e.g., Fabiani et al., 1987). For this electrode site, we find that the consistency and reliability of P300 depends on the way it is measured. An RMS amplitude measure provided the greatest reliability between blocks (test-retest) and consistency among laboratories. Peak amplitude tended to be reliable between blocks, but inconsistent among laboratories. Peak latency tended to be consistent among laboratories, but unreliable between blocks.

Our findings of superior reliability for an RMS measure question its continued low use, as noted in the literature. For example, Fabiani et al. (1987) observed that, between 1983 and 1987, 15 of 34 P300 studies exclusively used baseline-to-peak measures whereas only 4 exclusively used an area measure. The RMS is an area measure that is insensitive to polarity changes in the analysis interval. Fabiani et al. found that baseline-to-peak and area-amplitude measures in an auditory oddball task similar to ours showed about equal test-retest reliability ($r = .81$ or .80 respectively, $n = 50$ subjects). These reliability estimates are close to those obtained by NPRDC (.81 and .75); however, the other three laboratories obtained substantially lower baseline-to-peak amplitude reliabilities (see Table 13). The low peak amplitude reliabilities seen in these laboratories may be related to the low peak latency reliabilities they obtained. Peak latency differences between test and retest will affect peak amplitude measurements if the true peak moves outside of the measurement interval.

Fabiani et al. also found that P300 peak latency was less reliable than baseline-to-peak amplitude or an area measure similar to our RMS measure. They reported an r-value of .56, which compares closely with the present value obtained by NPRDC (.58) but is higher than that obtained by the other three laboratories.

Polich (1987) reported reliability measures of P300 in an auditory oddball paradigm that are somewhat lower than those obtained by Fabiani et al. but are comparable to some of the values we obtained. In an active discrimination condition, where subjects responded with a finger movement for rare tones, at site Pz, Polich found a test-retest correlation of 0.77 for P300 baseline-to-peak amplitude. This falls in the range of significant correlations obtained in this study (0.49 to 0.81). Polich's peak latency test-retest correlation of 0.26 was also lower than for amplitude. However, this value falls outside the range of significant correlations obtained in this study (0.42 to 0.58).

Overall, the reliability estimates from the literature clearly support superiority of amplitude versus latency measures of P300 in the auditory oddball task. The present data extend these results by demonstrating that an area measure of amplitude, the RMS, provides greater interlaboratory consistency than a baseline-to-peak amplitude measure.

At electrode sites other than Pz, the reliability of P300 was lower. In particular, at the frontal site (Fz), the data collected by NHRC differed from that collected by the other three laboratories. NHRC's average P300 amplitude at Fz was larger than that measured at the other laboratories and was significantly lower in the second block of trials than in the first block.

The reason for these differences is unclear; however, one possible explanation for this result is that the NHRC subjects had a high level of sonar monitoring experience, which involves extensive practice in performing auditory signal detection and discrimination. Other studies have shown that the distribution of ERP component amplitudes across the head may vary with experience level on a specific task. In particular, ERP amplitudes differ between experienced and naive subjects at site Fz (Kobus, Beeler, & Stashower, 1987; Kobus & Stashower, 1988). Age differences between the NHRC subjects and the other three laboratories (see Methods section) may also account for the test-retest differences at site Fz in the NHRC data. However, at NHRC, some recent (Merrill, 1990) data support the notion that effects on ERP amplitudes at site Fz due to experience differences are separable from those due to age differences.

Research has shown that P300 latency correlates with RT in tasks that emphasize response accuracy (discussed in Hillyard & Picton, 1986). The present study, which also stressed response accuracy, provides new estimates of the correlation between average P300 measures and average RT for the classification response to rare or frequent auditory stimuli. Under the conditions of our study, at site Pz, the correlation between P300 latency and RT means for hits and correct rejections was low and generally nonsignificant (Table 14). Correlations between baseline-to-peak amplitude at site Pz and RT were also low and nonsignificant. On the other hand, correlations between RMS amplitude at site Pz and RT were significant in two laboratories (NPRDC and NAMRL) and when pooled across laboratories. These coefficients were always negative, indicating that faster performance of the oddball task was indexed by larger P300 RMS amplitudes. However, some of this effect may be accounted for by the age differences in our samples, and their association with increasing RT and decreasing P300 amplitude (Picton, Stapells, Perrault, Baribeau-Braun, and Stuss, 1984).

As it was for the test-retest and interlaboratory comparisons of P300, the RMS measure of the oddball effect was also superior to baseline-to-peak amplitude and latency measures as an index of task performance. Its robustness may derive in part from its insensitivity to the variance associated with a single sample point in time, which is inherent in the baseline-to-peak and latency measures. As more points from the average ERP waveform are incorporated in an estimate, the expected value of the estimate will stabilize due to summation and cancellation of the random variances of individual sample points.

REFERENCES

Dixon, W. J. (Ed.). (1985). *BMDP statistical software*. Berkeley: University of California Press.

Fabiani, M., Gratton, G., Karis, D., & Donchin, E. (1987). Definition, identification, and reliability of measurement of the P300 component of the event-related brain potential. *Advances in Psychophysiology*, 2, 1-78.

Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, 55, 468-484.

Hillyard, S. A, & Picton, T. W. (1987). Electrophysiology of cognition. In S. A. Hillyard & T. W. Picton, *Handbook of Physiology, Section 1: The Nervous System, Volume V: Higher Functions of the Brain*, Part 2, Chapter 13. Bethesda, MD: American Physiological Society.

Jasper, H. (1958). The ten-twenty electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology*, 43, 397-403.

Kobus, D. A., Beeler, M. J., & Stashower, K. (1987). *Electrophysiological effects of experience during an auditory task* (NHRC Technical Report 87-2). San Diego: Naval Health Research Center.

Kobus, D. A, & Stashower, K. (1988). *Effects of experience and task difficulty on event-related potentials* (NHRC Technical Report 88-30). San Diego: Naval Health Research Center.

Merrill, L. (1990). *The effect of sonar experience on the event-related brain potential* (Unpublished doctoral dissertation). San Diego: United States International University.

Picton, T. W., Stapells, D. R., Perrault, N., Baribeau-Braun, J., & Stuss, D. T. (1984). Human event-related potentials: Current perspectives. In R. H. Nodar and C. Barber (Eds.), *Evoked Potentials II. The Second International Evoked Potentials Symposium*. Boston: Butterworth.

Polich, J. (1987). Comparison of P300 from a passive tone sequence paradigm and an active discrimination task. *Psychophysiology*, 24, 41-46.

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked potential correlates of stimulus uncertainty. *Science*, 150, 1187-1188.

Trejo, L. J. (1988, February). *RMS amplitude computations in ERP time windows* (NPRDC Letter 3900 Ser. 41/85). San Diego: Navy Personnel Research and Development Center.

APPENDIX

ERP COMPONENT MEASURES AND CORRELATIONS WITH PERFORMANCE/AGE DATA

ERP COMPONENT MEASURES AND CORRELATIONS WITH PERFORMANCE/AGE DATA

The following tables, A-1 through A-10, list means, standard deviations (S.D.), and correlation coefficients with performance and age variables of the ERP component measures reported by the laboratories, N1 (amplitude and latency), P2 (amplitude and latency, not reported by NHRC and UCSD), and P300 (amplitude, latency and RMS). Descriptions of these measures appear in the Methods section. The "Blk/Stim/Site" codes in each table stand for blocks (1,2), stimuli (F = frequent, R = rare), and sites (Fz, Cz, Pz). The codes for the performance variables are as follows: RTTF is reaction time in test session (block 1) for frequent stimuli. RTTR is reaction time in test session for rare stimuli. RTRF is reaction time in retest session (block 2) for frequent stimuli. RTRR is reaction time in test session for rare stimuli. PCT is percent correct in test sesssion; PCR is percent correct in retest session. AGE is calendar age of subject as assessed by self report. Degrees of freedom for each laboratory and corresponding critical values of the correlation coefficient at a significance level of $p < 0.05$ are:

NPRDC: df = 23, r = 0.398

NAMRL: df = 16, r = 0.468

NHRC: df = 6, r = 0.707

UCSD: df = 8, r = 0.632

Table A-1

NPRDC N1 Component Summary

Blk/Stim/Site	Statistics Mean	S.D.	Coefficients of Correlation With Performance Variables RTTF	RTTR	RTRF	RTRR	PCT	PCR	AGE
Summary Statistics and Correlations for N1 Peak Amplitude									
1/F/Fz	-7.61	2.65	0.37	0.30	0.22	0.26	0.11	-0.15	-0.07
1/F/Cz	-6.97	2.74	0.23	0.21	0.31	0.26	-0.01	-0.21	-0.23
1/F/Pz	-3.73	2.20	-0.03	0.01	0.26	0.12	0.01	-0.16	-0.25
1/R/Fz	-10.41	4.06	0.23	0.05	-0.02	0.03	-0.10	-0.31	-0.17
1/R/Cz	-10.47	4.47	0.07	-0.15	-0.14	-0.14	-0.18	-0.25	-0.16
1/R/Pz	-5.90	2.76	-0.04	-0.13	-0.10	-0.08	-0.25	-0.25	-0.18
2/F/Fz	-6.69	2.98	0.42	0.36	0.23	0.32	-0.26	-0.49	-0.10
2/F/Cz	-6.45	3.00	0.14	0.03	0.01	0.04	-0.41	-0.43	-0.08
2/F/Pz	-3.45	1.81	-0.09	-0.13	0.03	-0.02	-0.40	-0.24	-0.12
2/R/Fz	-10.13	4.86	0.43	0.34	0.29	0.39	-0.44	-0.73	-0.20
2/R/Cz	-10.61	4.52	0.38	0.27	0.25	0.32	-0.41	-0.62	-0.09
2/R/Pz	-6.25	2.75	0.24	0.08	0.07	0.11	-0.31	-0.48	-0.28
Summary Statistics and Correlations for N1 Peak Latency									
1/F/Fz	117.19	15.13	-0.15	-0.04	0.04	-0.07	-0.37	-0.05	-0.22
1/F/Cz	112.81	13.92	-0.20	0.05	0.17	0.07	-0.54	-0.27	-0.11
1/F/Pz	104.06	18.56	-0.23	0.00	-0.04	-0.06	-0.25	-0.05	0.08
1/R/Fz	120.31	15.30	-0.00	0.28	0.03	0.20	-0.21	-0.22	-0.13
1/R/Cz	118.13	17.52	-0.02	0.39	0.14	0.34	-0.20	-0.21	-0.00
1/R/Pz	114.69	21.95	0.05	0.39	0.09	0.30	-0.02	-0.06	0.07
2/F/Fz	116.88	16.80	-0.02	0.08	0.16	-0.01	-0.24	0.12	0.01
2/F/Cz	113.44	14.81	0.08	0.22	0.39	0.21	-0.42	-0.27	0.03
2/F/Pz	103.44	16.77	-0.08	0.07	0.10	0.05	-0.12	-0.06	0.12
2/R/Fz	118.44	17.86	-0.14	0.01	0.06	-0.05	-0.06	0.11	-0.06
2/R/Cz	116.25	16.00	-0.18	0.10	0.08	0.05	-0.14	0.02	-0.01
2/R/Pz	112.19	16.40	-0.30	-0.11	0.03	-0.09	-0.25	0.06	-0.05

Table A-2

NPRDC P2 Component Summary

Blk/Stim/Site	Statistics Mean	S.D.	Coefficients of Correlation With Performance Variables RTTF	RTTR	RTRF	RTRR	PCT	PCR	AGE
Summary Statistics and Correlations for P2 Peak Amplitude									
1/F/Fz	-2.21	3.16	0.08	-0.12	-0.20	-0.20	0.22	0.08	0.21
1/F/Cz	0.98	3.67	-0.00	-0.27	-0.20	-0.30	0.09	0.00	-0.10
1/F/Pz	3.20	4.34	-0.11	-0.31	-0.16	-0.29	0.14	0.04	-0.26
1/R/Fz	-4.26	5.51	0.20	-0.07	-0.11	-0.19	0.15	0.03	0.06
1/R/Cz	-2.45	7.79	0.11	-0.20	-0.19	-0.30	0.12	0.05	-0.08
1/R/Pz	0.98	5.44	0.01	-0.22	-0.22	-0.28	0.03	-0.01	-0.24
2/F/Fz	-1.63	2.99	0.07	-0.09	-0.13	-0.02	-0.15	-0.33	0.13
2/F/Cz	1.37	4.01	-0.17	-0.40	-0.43	-0.37	0.01	-0.04	-0.11
2/F/Pz	3.04	3.84	-0.28	-0.38	-0.41	-0.33	0.11	0.06	-0.20
2/R/Fz	-3.44	5.31	0.23	0.12	0.01	0.13	-0.18	-0.47	0.24
2/R/Cz	-1.17	6.74	0.16	-0.01	-0.03	-0.01	-0.08	-0.26	0.11
2/R/Pz	1.61	4.93	0.06	-0.17	-0.08	-0.12	-0.08	-0.16	-0.18
Summary Statistics and Correlations for P2 Peak Latency									
1/F/Fz	193.44	15.51	0.18	0.13	0.28	0.19	-0.20	-0.32	0.04
1/F/Cz	194.07	15.90	0.05	0.04	0.18	0.06	-0.20	-0.24	0.11
1/F/Pz	193.44	17.37	-0.19	-0.16	-0.05	-0.10	-0.22	-0.25	0.17
1/R/Fz	183.13	18.33	-0.00	0.02	-0.12	-0.13	0.44	0.40	0.06
1/R/Cz	190.32	17.01	0.20	0.04	0.12	0.00	0.09	0.03	-0.27
1/R/Pz	194.07	17.57	-0.05	0.14	0.24	0.17	-0.14	-0.05	-0.19
2/F/Fz	192.50	16.40	-0.41	-0.44	-0.26	-0.37	0.02	0.23	0.34
2/F/Cz	196.25	13.96	-0.57	-0.41	-0.39	-0.38	-0.09	0.20	0.25
2/F/Pz	195.63	14.17	-0.51	-0.23	-0.35	-0.26	-0.13	0.17	0.16
2/R/Fz	183.44	17.63	0.11	0.15	0.02	0.09	0.22	0.03	0.06
2/R/Cz	188.44	15.52	0.19	0.20	0.31	0.20	-0.34	-0.23	-0.10

Table A-3

NPRDC P300 Component Summary

Blk/Stim/Site	Statistics Mean	S.D.	Coefficients of Correlation With Performance Variables RTTF	RTTR	RTRF	RTRR	PCT	PCR	AGE
Summary Statistics and Correlations for P300 Peak Amplitude									
1/F/Fz	-0.00	3.29	0.12	-0.05	0.04	-0.06	-0.20	0.06	-0.03
1/F/Cz	1.94	4.71	-0.03	-0.25	0.02	-0.18	-0.15	-0.06	-0.10
1/F/Pz	4.84	5.14	-0.09	-0.28	-0.07	-0.24	0.03	0.01	-0.22
1/R/Fz	7.55	6.65	-0.43	-0.39	-0.38	-0.43	0.24	0.47	0.44
1/R/Cz	15.01	8.54	-0.37	-0.37	-0.31	-0.35	0.16	0.28	0.11
1/R/Pz	19.60	8.55	-0.25	-0.20	-0.15	-0.17	0.09	0.20	-0.01
2/F/Fz	-0.24	3.55	0.00	-0.10	-0.17	-0.10	-0.19	-0.20	-0.12
2/F/Cz	2.03	6.15	-0.10	-0.22	-0.25	-0.23	-0.12	-0.18	-0.22
2/F/Pz	5.21	5.79	-0.11	-0.23	-0.21	-0.22	0.01	0.01	-0.30
2/R/Fz	6.37	6.07	-0.51	-0.48	-0.42	-0.48	-0.02	0.41	0.29
2/R/Cz	14.30	9.01	-0.36	-0.41	-0.38	-0.42	0.12	0.38	0.01
2/R/Pz	19.50	8.77	-0.31	-0.32	-0.30	-0.32	0.13	0.28	-0.15
Summary Statistics and Correlations for P300 Peak Latency									
1/F/Fz	360.00	61.01	-0.01	-0.02	-0.16	-0.06	-0.15	-0.30	-0.58
1/F/Cz	350.00	63.75	0.12	0.18	0.07	0.17	-0.16	-0.36	-0.53
1/F/Pz	305.00	44.96	0.17	0.11	-0.06	-0.06	-0.18	-0.01	-0.05
1/R/Fz	356.88	23.08	0.65	0.70	0.53	0.54	-0.20	-0.46	-0.24
1/R/Cz	348.75	29.48	0.42	0.55	0.40	0.41	-0.13	-0.34	-0.13
1/R/Pz	355.94	26.88	0.45	0.59	0.33	0.39	0.06	-0.18	-0.12
2/F/Fz	365.00	63.65	0.02	0.06	-0.25	-0.06	0.29	0.09	-0.00
2/F/Cz	348.75	67.28	0.12	0.06	-0.23	-0.09	0.13	-0.05	-0.07
2/F/Pz	306 88	44.91	0.47	0.35	0.28	0.25	0.01	-0.17	-0.22
2/R/Fz	359.69	28.30	0.64	0.75	0.63	0.68	-0.14	-0.33	-0.05
2/R/Cz	352.81	30.73	0.40	0.60	0.53	0.59	-0.18	-0.36	-0.03
2/R/Pz	360.63	34.11	0.41	0.61	0.57	0.65	-0.18	-0.48	-0.07
Summary Statistics and Correlations for P300 RMS Amplitude									
1/F/Fz	3.36	1.92	-0.24	-0.01	-0.19	0.03	0.08	0.02	-0.17
1/F/Cz	3.49	2.00	-0.13	-0.14	-0.39	-0.25	0.19	0.20	-0.24
1/F/Pz	3.95	2.84	-0.10	-0.37	-0.30	-0.36	0.07	0.01	-0.32
1/R/Fz	6.63	3.33	-0.08	-0.04	-0.09	-0.03	0.25	0.08	0.41
1/R/Cz	11.06	5.52	-0.34	-0.44	-0.29	-0.41	0.16	0.33	0.12
1/R/Pz	13.99	6.37	-0.38	-0.39	-0.31	-0.34	0.03	0.18	0.01
2/F/Fz	3.15	1.84	-0.05	0.14	0.01	0.11	0.01	0.21	0.10
2/F/Cz	4.29	2.63	0.08	0.03	-0.00	-0.01	0.02	0.18	-0.21
2/F/Pz	4.05	3.84	-0.07	-0.23	-0.26	-0.25	0.04	0.06	-0.34
2/R/Fz	6.20	2.93	0.06	0.03	-0.01	0.08	-0.04	-0.09	0.14
2/R/Cz	10.87	5.44	-0.20	-0.36	-0.33	-0.35	0.08	0.18	-0.10
2/R/Pz	13.35	7.01	-0.42	-0.53	-0.47	-0.52	0.10	0.31	-0.14

Table A-4

NA. RL N1 Component Summary

Blk/S.n/Site	Statistics Mean	S D.	Coefficients of Correlation With Performance Variables RTTT	RTTR	RTRF	RTRR	PCT	PCR	AGE
Summary Statistics and Correlations for N1 Peak Amplitude									
1/F/Fz	-5.64	2.91	0.46	0.51	0.46	0.44	-0.03	0.04	0.07
1/F/Cz	-5.26	2.79	0.50	0.49	0.45	0.46	0.09	0.04	-0.08
1/F/Pz	-3.69	2.00	0.45	0.38	0.47	0.40	-0.07	-0.08	0.09
1/R/Fz	-7.95	3.33	0.11	0.12	0.12	0.23	0.01	0.12	-0.16
1/R/Cz	-8.32	3.97	0.17	0.18	0.18	0.29	0.04	0.08	-0.03
1/R/Pz	-5.84	3.16	0.01	0.02	0.04	0.07	0.13	0.06	-0.08
2/F/Fz	-4.9[illegible]	2.80	0.01	0.08	-0.08	0.06	0.12	0.29	-0.07
2/F/Cz	-4.87	3.3[illegible]	0.10	0.08	0.06	0.07	0.33	0.32	0.02
2/F/Pz	-3.46	2.76	0.25	0.06	0.21	0.10	0.31	0.08	0.06
2/R/Fz	-8.24	4.53	0.24	0.28	0.24	0.31	0.00	0.18	-0.30
2/R/Cz	-8.91	4.60	0.23	0.31	0.32	0.41	0.04	0.25	-0.19
2/R/Pz	-6.46	3.40	0.03	0.10	0.11	0.18	0.13	0.32	-0.29
Summary Statistics and Correlations for N1 Peak Latency									
1/F/Fz	95.56	14.83	-0.12	-0.13	-0.01	-0.26	0.12	0.15	0.22
1/F/Cz	102.22	21.04	-0.23	-0.12	-0.14	-0.30	0.24	0.33	0.21
1/F/Pz	102.89	23.07	-0.49	-0.28	-0.41	-0.38	-0.26	0.01	0.33
1/R/Fz	107.11	19.23	0.15	0.04	0.12	0.02	0.41	0.17	-0.17
1/R/Cz	106.44	18.62	0.11	0.12	0.09	-0.02	0.33	0.14	-0.11
1/R/Pz	98.44	19.51	0.32	0.40	0.27	0.32	0.16	0.18	-0.34
2/F/Fz	94.67	16.86	-0.17	-0.14	-0.10	-0.25	0.08	-0.05	0.20
2/F/Cz	99.33	22.05	-0.14	-0.15	-0.19	-0.21	0.08	0.01	0.14
2/F/Pz	92.44	22.12	-0.30	-0.38	-0.33	-0.43	0.3[illegible]	0.36	-0.06
2/R/Fz	103.11	16.72	-0.15	-0.09	0.00	-0.05	0.32	0.39	-0.27
2/R/Cz	107.78	18.28	-0.37	-0.19	-0.18	-0.21	-0.07	0.15	-0.16
2/R/Pz	94.89	18.40	-0.15	-0.11	-0.26	-0.21	0.32	0.48	0.06

Table A-5

NAMRL P2 Component Summary

Blk/Stim/Site	Statistics Mean	S.D.	Coefficients of Correlation With Performance Variables RTTF	RTTR	RTRF	RTRR	PCT	PCR	AGE
Summary Statistics and Correlations for P2 Peak Amplitude									
1/F/Fz	0.92	2.80	0.04	0.06	0.09	0.03	-0.06	-0.29	0.17
1/F/Cz	2.88	3.72	-0.17	-0.15	-0.16	-0.19	0.16	-0.13	0.01
1/F/Pz	3.59	3.52	-0.24	-0.24	-0.18	-0.25	-0.06	-0.21	0.01
1/R/Fz	0.19	3.19	0.25	0.11	0.21	0.18	-0.09	-0.33	0.03
1/R/Cz	1.31	4.51	0.34	0.25	0.27	0.29	0.03	-0.22	-0.02
1/R/Pz	2.56	3.38	0.32	0.12	0.30	0.15	0.30	-0.06	-0.07
2/F/Fz	1.04	3.06	-0.56	-0.40	-0.60	-0.40	0.04	0.12	-0.08
2/F/Cz	3.18	4.05	-0.48	-0.40	-0.54	-0.41	0.29	0.27	-0.11
2/F/Pz	3.28	3.22	-0.17	-0.20	-0.19	-0.17	0.49	0.42	-0.08
2/R/Fz	0.79	3.84	0.34	0.20	0.19	0.12	0.18	-0.14	-0.17
2/R/Cz	1.62	5.55	0.35	0.18	0.21	0.15	0.33	0.05	-0.19
2/R/Pz	2.45	4.37	0.20	0.06	0.15	0.04	0.40	0.15	-0.25
Summary Statistics and Correlations for P2 Peak Latency									
1/F/Fz	176.22	20.83	-0.46	-0.41	-0.62	-0.50	-0.10	-0.11	0.02
1/F/Cz	177.11	21.16	-0.44	-0.41	-0.62	-0.50	-0.05	-0.10	0.03
1/F/Pz	178.67	19.64	0.01	0.15	-0.03	0.07	-0.23	-0.10	-0.04
1/R/Fz	184.22	18.38	0.29	0.31	0.28	0.29	-0.22	-0.23	-0.36
1/R/Cz	184.00	16.69	0.47	0.34	0.41	0.38	-0.14	-0.22	-0.52
1/R/Pz	184.22	18.89	0.52	0.32	0.42	0.34	0.51	0.29	-0.44
2/F/Fz	180.22	20.14	-0.00	-0.02	-0.24	-0.25	0.08	-0.12	0.08
2/F/Cz	183.33	16.48	0.21	0.17	0.02	-0.11	0.35	-0.02	-0.34
2/F/Pz	178.67	16.97	-0.05	0.07	-0.12	-0.16	0.25	0.22	-0.23
2/R/Fz	182.22	17.95	0.29	0.15	0.22	0.01	0.53	0.07	0.38
2/R/Cz	180.44	17.99	0.43	0.26	0.33	0.09	0.44	0.00	0.33
2/R/Pz	184.44	16.23	0.42	0.34	0.42	0.32	0.23	0.13	0.28

Table A-6

NAMRL P300 Component Summary

	Statistics		Coefficients of Correlation With Performance Variables						
Blk/Stim/Site	Mean	S.D.	RTTF	RTTR	RTRF	RTRR	PCT	PCR	AGE
Summary Statistics and Correlations for P300 Peak Amplitude									
1/F/Fz	1.19	2.40	0.22	0.15	0.12	0.04	-0.09	-0.17	0.10
1/F/Cz	3.23	2.20	-0.15	-0.19	-0.26	-0.36	-0.18	-0.44	-0.01
1/F/Pz	4.12	2.68	-0.22	-0.22	-0.21	-0.26	-0.09	-0.16	-0.27
1/R/Fz	6.85	3.13	-0.40	-0.20	-0.42	-0.14	-0.31	-0.04	-0.03
1/R/Cz	11.85	5.79	-0.50	-0.34	-0.59	-0.41	-0.14	-0.15	-0.16
1/R/Pz	12.64	5.94	-0.28	-0.31	-0.35	-0.33	0.10	-0.02	-0.29
2/F/Fz	2.43	2.93	-0.46	-0.47	-0.49	-0.52	0.15	0.14	0.21
2/F/Cz	4.19	3.75	-0.47	-0.54	-0.47	-0.61	0.23	0.05	0.16
2/F/Pz	4.59	3.94	-0.22	-0.43	-0.18	-0.40	0.40	0.20	-0.01
2/R/Fz	6.36	4.17	-0.28	-0.05	-0.34	-0.10	0.02	0.38	-0.19
2/R/Cz	11.84	5.27	-0.54	-0.30	-0.66	-0.41	-0.00	0.17	-0.26
2/R/Pz	12.55	4.33	-0.42	-0.38	-0.54	-0.42	0.12	0.13	-0.39
Summary Statistics and Correlations for P300 Peak Latency									
1/F/Fz	341.33	48.16	-0.06	0.13	-0.07	0.05	-0.31	-0.22	0.03
1/F/Cz	335.56	46.87	0.10	0.19	0.24	0.15	-0.39	-0.13	0.44
1/F/Pz	323.78	46.21	0.27	0.33	0.42	0.24	-0.27	-0.17	0.51
1/R/Fz	347.11	31.96	0.38	0.37	0.43	0.45	-0.41	-0.29	-0.19
1/R/Cz	348.89	29.79	0.40	0.36	0.48	0.47	-0.40	-0.27	0.04
1/R/Pz	359.11	29.57	0.39	0.34	0.47	0.48	-0.34	-0.17	0.27
2/F/Fz	346.89	34.24	0.07	-0.01	-0.05	-0.11	0.13	0.04	-0.28
2/F/Cz	316.22	35.95	-0.43	-0.41	-0.30	-0.35	-0.31	-0.05	0.10
2/F/Pz	293.33	27.30	0.10	-0.03	0.24	0.00	0.08	0.07	-0.19
2/R/Fz	346.44	35.96	0.58	0.48	0.61	0.42	-0.25	-0.56	0.17
2/R/Cz	344.89	36.73	0.61	0.51	0.68	0.47	-0.21	-0.47	0.36
2/R/Pz	357.11	34.07	0.49	0.53	0.65	0.67	-0.54	-0.47	0.21
Summary Statistics and Correlations for P300 RMS Amplitude									
1/F/Fz	2.06	1.02	-0.03	-0.22	-0.08	-0.11	0.06	0.04	0.29
1/F/Cz	3.06	1.52	-0.22	-0.20	-0.24	-0.30	-0.16	-0.40	0.00
1/F/Pz	3.53	1.58	-0.27	-0.15	-0.28	-0.18	-0.21	-0.21	-0.35
1/R/Fz	4.97	2.44	-0.33	-0.15	-0.31	-0.14	-0.19	0.08	0.06
1/R/Cz	9.70	5.27	-0.51	-0.35	-0.57	-0.44	-0.01	-0.02	-0.15
1/R/Pz	10.81	5.27	-0.29	-0.32	-0.38	-0.39	0.22	0.09	-0.32
2/F/Fz	2.87	1.38	-0.00	-0.19	0.04	-0.19	0.09	0.13	0.29
2/F/Cz	3.67	2.14	0.02	-0.16	0.09	-0.26	0.37	0.21	0.23
2/F/Pz	3.44	2.27	-0.05	-0.23	-0.02	-0.29	0.50	0.26	0.05
2/R/Fz	5.37	2.82	-0.44	-0.25	-0.41	-0.32	0.02	0.24	-0.24
2/R/Cz	9.85	5.38	-0.63	-0.38	-0.70	-0.45	-0.08	0.13	-0.23
2/R/Pz	10.85	5.09	-0.53	-0.45	-0.62	-0.46	0.08	0.16	-0.41

Table A-7

NHRC N1 Component Summary

Blk/Stim/Site	Statistics Mean	S.D.	Coefficients of Correlation With Performance Variables RTTF	RTTR	RTRF	RTRR	PCT	PCR	AGE
Summary Statistics and Correlations for N1 Peak Amplitude									
1/F/Fz	-12.85	3.47	-0.04	0.74	0.54	0.65	-0.45	-0.74	-0.14
1/F/Cz	-9.78	3.96	0.01	0.76	0.45	0.69	-0.32	-0.53	-0.45
1/F/Pz	-5.13	3.80	0.13	0.73	0.34	0.71	-0.39	-0.60	-0.33
1/R/Fz	-18.91	3.58	0.26	0.46	0.50	0.49	-0.56	-0.47	-0.01
1/R/Cz	-15.38	5.34	0.31	0.79	0.51	0.80	-0.78	-0.83	-0.14
1/R/Pz	-8.38	3.70	0.19	0.82	0.37	0.83	-0.64	-0.83	-0.18
2/F/Fz	-14.53	4.29	-0.46	0.09	0.03	0.01	0.14	-0.27	0.39
2/F/Cz	-11.74	3.78	-0.38	0.44	0.11	0.33	-0.03	-0.56	0.15
2/F/Pz	-5.83	3.33	-0.16	0.59	0.09	0.53	-0.23	-0.71	0.07
2/R/Fz	-18.57	7.50	0.37	0.56	0.62	0.65	-0.63	-0.68	-0.03
2/R/Cz	-18.57	7.23	0.26	0.60	0.25	0.69	-0.50	-0.67	-0.06
2/R/Pz	-11.43	6.33	0.24	0.48	0.11	0.56	-0.40	-0.56	-0.00
Summary Statistics and Correlations for N1 Peak Latency									
1/F/Fz	98.40	13.87	0.04	0.62	0.54	0.56	-0.44	-0.50	-0.19
1/F/Cz	106.60	19.63	-0.03	-0.01	0.14	-0.11	-0.14	-0.16	0.34
1/F/Pz	94.00	16.47	-0.02	0.56	0.35	0.52	-0.46	-0.60	0.07
1/R/Fz	95.60	8.79	0.34	0.26	0.24	0.33	-0.15	-0.23	-0.37
1/R/Cz	94.60	11.60	0.50	0.03	0.43	0.12	-0.42	-0.39	0.20
1/R/Pz	92.00	11.47	0.43	-0.31	0.18	-0.20	-0.14	-0.00	0.33
2/F/Fz	93.60	6.57	0.25	0.76	0.64	0.66	-0.62	-0.84	-0.18
2/F/Cz	93.40	6.50	0.08	-0.10	-0.12	-0.14	-0.10	-0.15	0.21
2/F/Pz	93.60	14.56	-0.07	-0.30	-0.02	-0.29	0.12	0.25	0.06
2/R/Fz	102.40	12.80	0.14	-0.12	0.26	-0.05	0.07	0.19	-0.22
2/R/Cz	109.40	15.15	0.01	-0.18	-0.11	-0.18	0.04	0.22	-0.13
2/R/Pz	112.80	22.33	-0.35	-0.39	-0.40	-0.39	0.20	-0.03	0.64

Table A-8

NHRC P300 Component Summary

Blk/Stim/Site	Statistics Mean	S.D.	Coefficients of Correlation With Performance Variables RTTF	RTTR	RTRF	RTRR	PCT	PCR	AGE
Summary Statistics and Correlations for P300 Peak Amplitude									
1/F/Fz	4.50	4.39	0.05	-0.13	0.09	-0.02	0.10	0.31	-0.10
1/F/Cz	3.36	4.94	-0.07	0.23	0.02	0.30	-0.05	0.02	-0.31
1/F/Pz	5.38	4.31	0.34	0.74	0.51	0.76	-0.53	-0.46	-0.72
1/R/Fz	16.76	7.67	-0.22	-0.88	-0.65	-0.78	0.78	0.89	0.28
1/R/Cz	13.38	6.57	-0.58	-0.66	-0.63	-0.72	0.75	0.80	0.15
1/R/Pz	16.80	7.15	-0.22	-0.02	-0.12	-0.15	0.24	0.39	-0.48
2/F/Fz	3.71	5.42	-0.06	-0.32	-0.14	-0.20	0.23	0.41	0.12
2/F/Cz	0.78	6.05	-0.23	0.13	-0.01	0.17	0.02	-0.09	0.07
2/F/Pz	3.05	3.46	0.03	0.49	0.24	0.51	-0.29	-0.21	-0.47
2/R/Fz	9.99	11.21	-0.46	-0.44	-0.18	-0.50	0.48	0.37	0.44
2/R/Cz	9.86	7.40	-0.41	-0.12	0.02	-0.20	0.15	0.06	0.32
2/R/Pz	16.37	6.31	-0.22	0.28	0.29	0.13	-0.01	0.02	-0.42
Summary Statistics and Correlations for P300 Peak Latency									
1/F/Fz	338.80	19.38	-0.06	0.54	0.35	0.43	-0.23	-0.62	-0.17
1/F/Cz	337.80	16.50	0.57	-0.10	0.43	0.03	-0.27	-0.11	0.10
1/F/Pz	320.80	28.28	0.31	0.38	0.41	0.53	-0.50	-0.49	-0.02
1/R/Fz	332.80	23.64	-0.47	-0.66	-0.48	-0.68	0.76	0.79	0.22
1/R/Cz	348.20	13.87	0.14	0.70	0.62	0.51	-0.54	-0.66	-0.31
1/R/Pz	356.00	13.41	0.55	0.43	0.80	0.44	-0.60	-0.32	-0.50
2/F/Fz	346.80	12.91	0.11	0.49	0.59	0.36	-0.31	-0.56	-0.18
2/F/Cz	339.40	30.07	0.30	0.43	0.67	0.40	-0.33	-0.45	-0.22
2/F/Pz	314.00	28.36	0.36	-0.01	0.41	0.07	-0.27	-0.14	0.20
2/R/Fz	335.20	32.35	0.11	-0.54	-0.20	-0.41	0.49	0.77	-0.19
2/R/Cz	350.20	21.50	0.55	0.35	0.36	0.44	-0.31	-0.05	-0.77
2/R/Pz	354.80	17.34	0.18	0.05	-0.06	0.14	-0.34	-0.34	0.22
Summary Statistics and Correlations for P300 RMS Amplitude									
1/F/Fz	4.42	1.66	0.31	-0.01	0.15	0.15	-0.18	0.30	-0.37
1/F/Cz	3.97	2.55	0.32	-0.15	0.39	-0.14	-0.09	0.18	0.00
1/F/Pz	4.41	2.80	0.65	0.75	0.81	0.78	-0.73	-0.56	-0.72
1/R/Fz	12.76	3.52	0.01	-0.81	-0.50	-0.67	0.61	0.79	0.19
1/R/Cz	10.04	4.13	-0.61	-0.41	-0.45	-0.54	0.59	0.59	0.05
1/R/Pz	11.57	5.94	-0.31	-0.16	-0.24	-0.28	0.41	0.56	-0.42
2/F/Fz	5.17	1.82	0.75	0.76	0.93	0.84	-0.95	-0.75	-0.42
2/F/Cz	4.94	3.63	-0.05	-0.20	0.09	-0.33	0.05	0.01	0.34
2/F/Pz	3.30	1.60	0.42	0.24	0.73	0.17	-0.44	-0.19	-0.24
2/R/Fz	11.19	4.70	-0.19	0.66	0.52	0.47	-0.40	-0.74	-0.01
2/R/Cz	8.28	4.02	-0.26	0.26	0.37	0.14	-0.10	-0.25	0.05
2/R/Pz	10.26	4.49	-0.39	0.07	0.07	-0.10	0.28	0.28	-0.39

Table A-9

UCSD N1 Component Summary

Blk/Stim/Site	Statistics Mean	S.D.	Coefficients of Correlation With Performance Variables RTTF	RTTR	RTRF	RTRR	PCT	PCR	AGE
Summary Statistics and Correlations for N1 Peak Amplitude									
1/F/Fz	-5.79	1.67	-0.78	-0.39	-0.63	-0.36	-0.65	-0.65	0.51
1/F/Cz	-4.34	2.08	-0.85	-0.58	-0.77	-0.56	-0.51	-0.50	-0.05
1/F/Pz	-2.27	1.30	-0.52	-0.08	-0.38	-0.17	-0.73	-0.71	0.25
1/R/Fz	-9.71	3.22	-0.42	-0.94	-0.41	-0.87	0.05	-0.16	0.29
1/R/Cz	-7.75	6.88	-0.29	-0.67	-0.37	-0.66	0.10	-0.28	0.19
1/R/Pz	-6.03	3.48	-0.42	-0.72	-0.45	-0.78	-0.07	-0.43	0.19
2/F/Fz	-5.45	1.52	-0.45	-0.16	-0.26	-0.13	-0.62	-0.46	0.31
2/F/Cz	-4.58	1.75	-0.68	-0.53	-0.56	-0.52	-0.56	-0.45	-0.13
2/F/Pz	-2.29	1.59	-0.59	-0.69	-0.49	-0.67	-0.35	-0.30	0.05
2/R/Fz	-9.48	2.63	-0.58	-0.50	-0.60	-0.41	0.12	0.16	0.18
2/R/Cz	-7.81	2.77	0.09	-0.33	0.02	-0.26	0.45	0.44	-0.44
2/R/Pz	-5.72	2.49	-0.50	-0.65	-0.53	-0.50	0.22	0.14	0.03
Summary Statistics and Correlations for N1 Peak Latency									
1/F/Fz	95.20	20.90	-0.46	-0.59	-0.30	-0.54	-0.58	-0.53	0.01
1/F/Cz	96.40	17.93	-0.51	-0.55	-0.37	-0.51	-0.56	-0.46	0.03
1/F/Pz	101.20	18.86	-0.59	-0.25	-0.43	-0.32	-0.87	-0.92	0.34
1/R/Fz	105.60	19.25	-0.60	-0.30	-0.48	-0.37	-0.69	-0.71	0.21
1/R/Cz	98.80	17.18	-0.43	-0.16	-0.31	-0.21	-0.59	-0.77	0.70
1/R/Pz	95.60	14.66	-0.68	-0.85	-0.67	-0.88	-0.28	-0.50	0.34
2/F/Fz	94.40	17.30	-0.53	-0.70	-0.39	-0.64	-0.53	-0.51	0.06
2/F/Cz	92.00	13.33	-0.63	-0.54	-0.65	-0.62	-0.27	-0.24	-0.16
2/F/Pz	100.00	19.04	-0.56	-0.53	-0.46	-0.49	-0.47	-0.34	-0.18
2/R/Fz	106.40	24.24	-0.56	-0.28	-0.41	-0.33	-0.76	-0.74	0.16
2/R/Cz	107.20	21.89	0.01	-0.05	0.19	-0.12	-0.67	-0.65	0.15
2/R/Pz	98.40	14.01	-0.72	-0.52	-0.61	-0.63	-0.74	-0.81	0.39

Table A-10

UCSD P300 Component Summary

	Statistics		Coefficients of Correlation With Performance Variables						
Blk/Stim/Site	Mean	S.D.	RTTF	RTTR	RTRF	RTRR	PCT	PCR	AGE
Summary Statistics and Correlations for P300 Peak Amplitude									
1/F/Fz	3.25	1.82	0.04	-0.41	0.12	-0.34	-0.04	0.04	-0.25
1/F/Cz	3.16	1.45	0.13	-0.50	0.14	-0.39	0.21	0.10	-0.06
1/F/Pz	2.95	0.98	0.40	-0.18	0.35	-0.11	0.53	0.53	0.10
1/R/Fz	10.52	7.03	-0.02	-0.05	-0.04	0.00	0.26	0.49	-0.27
1/R/Cz	14.49	6.40	-0.08	-0.14	-0.13	-0.07	0.31	0.50	-0.21
1/R/Pz	14.31	4.84	-0.11	-0.21	-0.14	-0.12	0.24	0.43	-0.05
2/F/Fz	3.01	1.68	0.02	-0.18	-0.07	-0.27	0.25	0.28	-0.47
2/F/Cz	2.74	1.99	-0.22	-0.80	-0.29	-0.83	0.09	-0.09	-0.16
2/F/Pz	3.23	1.57	-0.37	-0.88	-0.36	-0.86	0.01	-0.07	0.20
2/R/Fz	9.21	5.84	-0.01	-0.10	-0.11	-0.14	0.43	0.43	0.04
2/R/Cz	12.30	5.85	-0.09	-0.45	-0.19	-0.46	0.48	0.41	0.21
2/R/Pz	12.03	6.06	-0.16	-0.50	-0.21	-0.50	0.26	0.25	0.18
Summary Statistics and Correlations for P300 Peak Latency									
1/F/Fz	355.60	33.86	-0.17	-0.56	-0.22	-0.45	0.20	0.04	-0.22
1/F/Cz	349.20	34.26	0.04	0.23	0.11	0.43	0.14	0.33	0.10
1/F/Pz	323.60	37.59	0.09	-0.13	-0.03	-0.04	0.68	0.57	0.02
1/R/Fz	345.60	26.81	0.01	0.68	0.09	0.61	-0.54	-0.39	-0.14
1/R/Cz	345.20	24.30	-0.23	0.47	-0.14	0.42	-0.52	-0.47	0.38
1/R/Pz	341.60	23.19	-0.22	0.28	-0.19	0.22	-0.27	-0.38	0.64
2/F/Fz	344.40	36.00	-0.13	-0.46	-0.20	-0.33	0.26	0.09	-0.31
2/F/Cz	336.40	27.87	-0.11	-0.46	-0.14	-0.30	0.22	0.18	-0.41
2/F/Pz	303.60	27.93	0.42	-0.12	0.33	-0.01	0.63	0.60	-0.08
2/R/Fz	332.80	33.46	0.06	0.23	-0.14	0.09	0.40	0.16	-0.05
2/R/Cz	340.40	17.83	-0.28	0.16	-0.23	0.15	-0.34	-0.53	0.29
2/R/Pz	350.80	25.51	-0.33	-0.18	-0.35	-0.11	0.09	-0.09	0.45
Summary Statistics and Correlations for P300 RMS Amplitude									
1/F/Fz	2.17	1.02	0.16	-0.02	0.11	-0.02	0.38	0.49	-0.39
1/F/Cz	2.42	0.79	0.21	-0.29	0.21	-0.13	0.30	0.43	-0.05
1/F/Pz	2.26	0.78	0.53	0.12	0.44	0.19	0.67	0.74	0.23
1/R/Fz	8.16	5.37	-0.02	0.10	-0.04	0.14	0.21	0.51	-0.19
1/R/Cz	11.11	5.33	-0.10	-0.20	-0.17	-0.12	0.37	0.56	-0.14
1/R/Pz	10.67	4.17	-0.11	-0.38	-0.16	-0.26	0.41	0.53	0.05
2/F/Fz	2.14	0.91	0.01	-0.02	-0.04	-0.13	0.18	0.14	-0.15
2/F/Cz	2.17	1.19	-0.42	-0.58	-0.42	-0.72	-0.33	-0.51	0.30
2/F/Pz	2.40	1.34	-0.39	-0.69	-0.31	-0.67	-0.16	-0.25	0.74
2/R/Fz	6.62	4.77	-0.16	-0.18	-0.22	-0.18	0.34	0.38	0.17
2/R/Cz	9.28	5.65	-0.22	-0.50	-0.31	-0.49	0.42	0.36	0.30
2/R/Pz	9.01	5.23	-0.21	-0.54	-0.26	-0.55	0.27	0.25	0.22

DISTRIBUTION LIST

Distribution:
Defense Technical Information Center (2)
Director, Office of Naval Research (OCNR-10)
Head, Cognitive and Neuro Sciences (OCNR-1142)
Life Sciences Directorate (OCNR-30T)
Cognitive and Decision Science (OCNR-1142CS)
Chief of Naval Research (ONT-20), (222)
Naval Reserve Officers Training Corps Division (Code N-1)
Chief of Naval Education and Training (L01)
Commander, Air Force Human Resources Laboratory (AFHRL/ID)
TSRL/Technical Library (FL 2870)
Director of Research, U.S. Naval Academy